Nor Jnl Ling 40.3, 287–311 © Nordic Association of Linguists 2017 doi:10.1017/S0332586517000075

Lindén, Krister, Heidi Haltia, Juha Luukkonen, Antti O. Laine, Henri Roivainen & Niina Väisänen. 2017. FinnFN 1.0: The Finnish frame semantic database. *Nordic Journal of Linguistics* 40(3), 287–311.

FinnFN 1.0: The Finnish frame semantic database

Krister Lindén, Heidi Haltia, Juha Luukkonen, Antti O. Laine, Henri Roivainen & Niina Väisänen

The article describes the process of creating a Finnish language FrameNet or FinnFN, based on the original English language FrameNet hosted at the International Computer Science Institute in Berkeley, California. We outline the goals and results relating to the FinnFN project and especially to the creation of the FinnFrame corpus. The main aim of the project was to test the universal applicability of frame semantics by annotating real Finnish using the same frames and annotation conventions as in the original Berkeley FrameNet project. From Finnish newspaper corpora, 40,721 sentences were automatically retrieved and manually annotated as example sentences evoking certain frames. This became the FinnFrame corpus. Applying the Berkeley FrameNet annotation conventions to the Finnish language required some modifications due to Finnish morphology, and a convention for annotating individual morphemes within words was introduced for phenomena such as compounding, comparatives and case endings. Various questions about cultural salience across the two languages arose during the project, but problematic situations occurred only in a few examples, which we also discuss in the article. The article shows that, barring a few minor instances, the universality hypothesis of frames is largely confirmed for languages as different as Finnish and English.

Keywords: agglutinative language, annotation conventions, corpus creation, Finnish FrameNet, semantic annotation, universal applicability of frames

Krister Lindén, Nykykielten laitos, PL 24, 00014 Helsingin yliopisto, Unioninkatu 40, Finland.
krister.linden@helsinki.fi, heidi.haltia@gmail.com, juha.luukkonen@iki.fi, antti.o.laine@helsinki.fi, hege.roivainen@helsinki.fi, niina.vaisanen@helsinki.fi

1. INTRODUCTION

This article illustrates the process of creating a Finnish language FrameNet, henceforth FinnFN, based on the original English language FrameNet hosted at the International Computer Science Institute in Berkeley, California (Baker, Fillmore & Lowe 1998, Fontenelle 2003), henceforth BFN, and the goals and results relating to the Finnish project and to the creation of the FinnFrame corpus.[1] The main aim of the project was to test the universal applicability of frame semantics by annotating Finnish using the BFN annotation conventions, which are outlined below.

CrossMark

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

FrameNet language	Number of frames	Number of lexical units	Number of annotated examples
Finnish FrameNet	938	6,639	40,721
Berkeley FrameNet	1,221	13,577	202,222
French FrameNet	105	1,100	16,000
Chinese FrameNet	323	3,947	18,000
Hebrew FrameNet	167	3,006	19,268
Swedish FrameNet	1,194	38,700	9,006

Table 1. Overview of the size of the the Finnish FrameNet compared with the Berkeley, Chinese, French, Hebrew and Swedish FrameNets.

Example sentences were automatically retrieved from Finnish newspaper corpora, and resulted in a total of 40,721 manually annotated sentences evoking one frame per sentence. In Berkeley FrameNet, collections of one frame per sentence are also referred to as Lexicographic Annotation Sets. This annotated collection of sentences became the FinnFrame corpus. Due to the structure of the Finnish language, applying the FrameNet annotation conventions to Finnish required some modifications, as Finnish needed a convention for annotating individual morphemes within words. Questions about cultural salience across the two languages arose during the project, but did not result in any major obstacles and for the vast majority of semantic frames established in the BFN project, it was fairly straightforward to find corresponding Finnish examples.

BFN is a database of annotated material organized around the principles of FRAME SEMANTICS, a theory of meaning developed by Charles J. Fillmore (Fillmore 1976, 1982), according to which all meaning is best understood as SEMANTIC FRAMES – descriptions of events or entities and the participants in relation to them. These frames are thought to be universal to a certain extent, at least within the western cultural sphere. This hypothesis of universality is what the FinnFN project set out to test.

In addition to English and Finnish, FrameNet-related projects and FrameNet-inspired research also exists in German (Burchardt et al. 2006), Spanish (Subirats & Petruck 2003), Swedish (Borin et al. 2010), Brazilian Portuguese (Duran & Aluísio 2011), Japanese (Ohara et al. 2004), Chinese (You & Liu 2005), Korean (Hahm et al. 2014), French (Meurs et al. 2008), Danish (Bick 2011), Polish (Zawisławska, Derwojedowa & Linde-Usiekniewicz 2008), Italian (Tonelli & Pianta 2008, Lenci, Johnson & Lapesa 2010), Slovenian (Lönneker-Rodman 2007a) and Hebrew (Hayoun & Elhadad 2015). For a list of FrameNets in other languages, see Appendix. For an overview of different practices in creating FrameNet resources in various languages, see Lönneker-Rodman (2007b). Table 1 presents a comparison of some FrameNet resources based on the FrameNet site[2] for BFN, French and Chinese, and on

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

presentations at the Language Resources and Evaluation Conference (LREC) 2016 for Swedish[3] and Hebrew.[4] The table shows that FinnFN is the second largest when looking at the number of annotated examples and third largest when looking at the number of frames and lexical units. See also Friberg Heppin & Friberg (2012) for a more comprehensive overview of FrameNet resources.

The rest of the article is structured as follows: Section 1 continues with an outline of Fillmore's theory and the BFN project. In Section 2, we discuss the creation of the FinnFN corpus and the annotation process. In Section 3, we consider annotation problems encountered in relation to the nature of the Finnish language and how we propose to solve them. In Section 4, we discuss the applicability of the original English language frames to Finnish data as well as some possibilities for further research and use of the data. In Section 5, we conclude the article.

1.1 Fillmore's frame semantics and FrameNet

The original FrameNet project is based on the theory of frame semantics developed by Charles J. Fillmore (Fillmore 1976, 1982). The basis of frame semantics within cognitive science is in the concept of COGNITIVE FRAMES by which language users interpret their experiences. Frame semantics is the study of how language users associate linguistic forms with these cognitive structures, i.e. FRAMES (Fillmore & Baker 2010:313–314).

Using frames is a cognitive process to interpret information. At the core of the theory of frame semantics is the analysis of these processes (Fillmore & Baker 2010:316). All content words, such as verbs, adjectives and nouns, evoke a frame within which their meaning is interpreted (Fillmore & Baker 2010:318). The FrameNet project compiles lexical units into background frames, which can be described in terms of the common schematic properties of the lexical units evoking the frame. Lexical units are considered to evoke the same frame when their use triggers the same cognitive schema, i.e. they express similar real world situations and phenomena. The requirement for evoking the same frame is the ability to use the lexical units as semantic paraphrases or near-paraphrases of each other. However, the lexical units do not have be in the same lexical category. It is their common cognitive frame that is key.

1.2 Theory and annotation conventions of BFN

The BFN project is aiming to create an online lexical resource for English (Ruppenhofer et al 2010:5). FrameNet focuses on analyzing and constructing semantic frames, for example situations, events, processes, states, physical and visual characteristics and a myriad of other phenomena. In theory, there can be a limitless number of different frames, and thus FrameNet will never be complete.

In FrameNet, the element that evokes a frame is called FRAME EVOKING ELEMENT (FEE) also known as a lexical unit. In the food cooking frame, the verb denoting the

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

act of preparing food will trigger the Cooking_creation frame. The semantic arguments and complements included in the event or phenomenon are called FRAME ELEMENTS (FE). The frame elements are labelled according to their semantic role or content. In the Cooking_creation frame, the frame elements include the cook and the food that is being prepared. In the example sentence *John fries eggs*, the actor, *John*, is the COOK, *fries* the FEE that evokes the frame, and *eggs* the end product, PRODUCED_FOOD.

Frame elements are also divided into defining CORE ELEMENTS and PERIPHERAL ELEMENTS according to their relevance to the phenomenon described in the frame. In the cooking frame, for example, while the COOK is essential, an element indicating the time of the event is general and not a defining element of the frame. Core elements vary greatly from frame to frame, whereas peripheral elements, for example, TIME and PLACE are more ubiquitous and therefore recurrent.

Frames vary greatly in their level of generality and abstractness, as well as in the number of frame elements they incorporate. Sometimes one frame has all the properties of another frame and elaborates on it in some way (Fillmore, Johnson & Petruck 2003:239). A more fine-grained subframe inherits from a semantically more general frame. The relations between all the frames in FrameNet form a frame hierarchy. The same sentence can often be annotated as an example of a more general frame as well as an example of a more specific frame lower in the hierarchy. For example, the frames Fleeing and Travel inherit from the more general frame of Self_motion, which in turn evokes the semantic sphere of Motion. The word class of the FEE can also be a defining feature. In BFN, the word class sometimes functions as the distinguishing criterion between frames, as in the case of the frames Sounds and Make_noise covering nouns and verbs, respectively. Similar to the frames themselves, specific formal criteria may also apply to frame elements. The element ROLE in the frame Judgment_communication is syntactically defined as follows:

> ROLE is used for the capacity in which the EVALUEE is judged, and is expressed in as-PPs.[5]

2. THE FINNISH LANGUAGE FRAMENET

FinnFN is entirely based on the frames outlined in BFN. In December 2015, there were 1,215 frames in BFN. New frames are created and divided into more detailed subcategories, and there will always be new points of views to consider. The primary objective of the FinnFN project was to test the hypothesis of language-independent frames and investigate how well the frames of BFN apply to Finnish. It was assumed that the frame elements defined by BFN would be applicable to Finnish as such.

To a certain extent, the frames are thought to be independent of specific languages within the western cultural sphere, and no frames tied specifically to Finnish cultural

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

phenomena have been created in the FinnFN project. This approach is unlike the ones used in other projects, such as the Swedish, German and French FrameNet projects, where new frames were created and the existing English-based frames were in some cases merged or divided to adapt the set of frames to the languages in question. For example, in the Japanese FrameNet, new frames were constructed when the annotators found a frame that was necessary for either the Japanese or both the Japanese and English lexicons (Ohara 2012:1559–1560). The French FrameNet constructed new frames and remodeled existing ones (Candito et al. 2014). The Saarbrücken Lexical Semantics Acquisition (SALSA) project constructed so-called proto-frames for senses of German words or multi-word expressions that could not be matched with the frames existing in version 1.3 of BFN. For instance, the German verb *gelten* has a sense 'be directed at', predicated of actions. This sense is somewhat similar to the `Aiming` frame but crucially the verb does not allow the expression of the (core) FE AGENT, precluding a straightforward inclusion of the verb in that frame. There currently is still no BFN frame available that this German lexical unit (LU) can be assigned to. However, this does not imply that there is no common shared frame for this phenomenon in English and German. It simply means that SALSA has not been actively maintained, so its final release may not reflect the progress in BFN.

In order to test the universality hypothesis of frames, the Finnish project adopted a frame-by-frame approach (detailed in Section 2.2 below) which followed the annotation system of the original English frames as closely as possible to test the cross-lingual applicability of BFN. The idea was also to gain time by not duplicating the many decades of lexicographical work invested in the original FrameNet effort getting a head start for the Finnish FrameNet while at the same time mastering the details of the original FrameNet to make informed extensions at a later stage.

2.1 Searching for Finnish language example sentences

For the FinnFN project, around 80,000 English sentences representing 866 frames were chosen from the FrameNet version 1.5 corpus[6] (https://framenet.icsi.berkeley.edu/), after which the annotated parts of the sentences were translated into Finnish by professional translators. The Finnish translations with their English original sentences constitute a separate corpus (the English–Finnish TransFrame corpus) which is further analyzed in a separate article (Lindén et al. to appear). The main finding of that article is that 92% of the frames were transferred with their meaning intact when the frame elements were translated from English to Finnish. However, in the current article, we focus on the frame universality hypothesis using attested natural language as our starting point. We also focus on the modifications needed in the FrameNet annotation conventions in order to annotate authentic Finnish example sentences.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

After translating the English FEEs and the FEs into Finnish, the Finnish translations were used in an automated search for similar sentences in the National Library Newspaper corpus,[7] a corpus of Finnish newspapers and magazines compiled by the National Library of Finland comprising five billion words of running text.[8] The text was preprocessed with the Turku Dependency Parser.[9] The corpus search was carried out in order to find attested natural language examples of frame evoking sentences to ensure that the translated frames were not mere 'translationese'. The goal was to find out whether there exist example sentences in Finnish for all the frames in BFN included in the parsed FrameNet 1.5. corpus. Because we sampled from authentic Finnish language, we could examine whether the frame elements defined by BFN also occurred in authentic Finnish sentences. In addition, it was possible to see if and how the syntax and semantics of the annotated Finnish sentences differed from English, for example, whether the annotated Finnish example sentences contained semantic roles that could not be covered by the frame elements of the original BFN frame.

In our attempt to find authentic Finnish example sentences, different search strategies were tried using Corpus Workbench queries in Korp (Borin, Forsberg & Roxendal 2012) in order to find which automatic searches gave the best results. The best search strategy (in which 36.7% of the search results were accepted for manual frame annotation) was to use the word form, lemma and the syntactic dependence label of the FEE. The second best search strategy (in which less than 30% of the search results were accepted for manual frame annotation) was to determine the word form, lemma and the syntactic dependence label of the FEE as well as the dependence label of the FEs immediately to the right or left of the FEE. The worst search strategy (only 0.1% good examples) was to narrow down the search too much by using the word form, lemma and dependence labels of both the FEs and the FEE. Only for 14 frames did we have to manually retrieve example sentences as the automatic search strategies did not provide any suitable results. These problematic frames are further discussed in Section 4.

The low success rate of some of the search strategies mainly indicates that although five billion words is a big corpus, it is not big enough for fully automating the procedure to find authentic examples. The results of the automatic search strategies were therefore further pruned manually. All the search results were read and the best examples were selected for annotation. An example was considered good if it evoked the same frame as the original English example and preferably included the same frame elements. The main focus was on the FEE and other core elements. The discarded sentences are relevant examples of ambiguous lexemes as FEEs, which are disambiguated into specific frames by the core FEs in the context. Some of the discarded examples sentences were collected for future use as they are relevant negative examples when training a machine to automatically identify examples of frames in running text.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

2.2 Annotating the FinnFN corpus

Each example sentence deemed to represent its intended frame was manually annotated with the corresponding FEE and FEs. The annotation method that was used in the FinnFN project can be described as a FRAME-BY-FRAME approach as opposed to the full-text annotation or the LEMMA-BY-LEMMA approach used for instance in the German SALSA project (Burchardt et al. 2006) and Japanese FrameNet project (Ohara et al. 2004, Ohara 2012). The frame-by-frame approach assumes that the sentences to be annotated contain a lexical unit evoking the intended frame, in contrast to the lemma-by-lemma approach, which takes a sentence and determines which frames its lemmas might evoke. If an FEE in a sentence context clearly evoked an unintended frame, the sentence was left unannotated. For more ambiguous cases, the adopted approach probably resulted in more varying FEEs for certain frames than the lemma-by-lemma approach would have yielded, as a large sample of sentences from different topics were evaluated. As stated above, we created no new frames at this point, instead the frame-by-frame approach made it easier to account for the diversity of FEEs that may evoke a frame. Because the FinnFN project used the frames and annotation conventions originally created for English, some special characteristics of the Finnish language became readily observable when applying the BFN annotation conventions to Finnish.

After an initial round of manual annotation, which was done by four linguists working in parallel on the automatically retrieved sentences representing all frames, a second round of annotation was carried out, in which one linguist checked all the sentences representing each frame. During this second round, with the experience of having annotated the full corpus, examples could be reassigned to more suitable frames, which resulted in example sentences representing 938 frames. Typically the examples would be assigned to more specific frames during the second round, as new frames had been added to BFN during the course of the FinnFN project.

The FinnFN corpus currently (June 2017) consists of 40,721 annotated example sentences (with approximately one million lexemes) that represent 938 frames. The FinnFrame corpus is thus one of the biggest FrameNet projects so far (together with the BFN and the Swedish and the German FrameNet projects) and can therefore readily be utilised in corpus-linguistics for frame semantic research. The applicability of the English frames and frame elements in Finnish will be further discussed in Section 4. In most cases, both the frames and the frame elements based on English apply to Finnish without modifications. Each FE and its use in the FinnFN corpus has been described in detail on the FinnFN wiki page.[10]

As in BFN, in addition to semantic criteria, morphosyntactic criteria may also apply to frame elements in the FinnFN annotation. For example, in Finnish, the frame element TOPIC is often expressed with the elative suffix (*-stA*), and MESSAGE

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

Layer 1	Layer 2	Layer 3	Corpus example	Gloss
			Sekä	both
			aitoja	real
			että	and
<FEE>	<entity		*epäaitoja*	fake
	entity>	<reference_category>	*koruja*	jewellery
			kannattaa	should
			hoitaa	be treated
			hyvin	well

Table 2. A corpus example *Sekä aitoja että epäaitoja koruja kannattaa hoitaa hyvin* 'Both real and fake jewellery should be treated well' with annotations for the frame `Artificiality` using *epäaitoja* 'fake.a' as FEE on Layer 1, the entity *epäaitoja koruja* 'ARTIFICIAL jewellery' on Layer 2 and *koruja* 'jewellery' as the reference category on Layer 3.

is typically a subordinate clause beginning with the conjunction *että* 'that'. This is exemplified in examples (1) and (2).

(1) Francis vihjaisi, että aiemmin IAAF ei KERTONUT
Francis hint.PST that previously IAAF NEG tell.PST.CONNEG
[TOPIC kaiki**sta** tapauksi**sta**]. (`Telling`)
all.ELA case.PL.ELA

(2) Francis VIHJAISI, [MESSAGE **että** aiemmin IAAF ei kertonut
Francis hint.PST that previously IAAF NEG tell.PST.CONNEG
kaikista tapauksista]. (`Communication`)
all.ELA case.PL.ELA
'Francis hinted that previously the IAAF did not tell about all cases.'

The examples above demonstrate the notational conventions we have assumed in this paper: FEEs in example sentences are written in capital letters and the intended frame is stated after the example sentence in parentheses in a monospaced font. Other frame elements, when relevant, are enclosed in square brackets, annotated with the element's label in small capitals. If no language is specified, the examples are in Finnish or English.

Table 2 provides an example of the annotation layout used in the FinnFN project. The table shows an example sentence annotated with the frame `Artificiality`, in which an entity is evaluated for whether it evokes a particular reference category. In our annotation work, we used a three-layered annotation layout with tags on separate layers in order to express all relevant frame elements for a word. The separate layers are numbered for clarity so that the FEE, is on Layer 1, other FEs on Layer 2 and, when frame elements overlap, additional FEs are on Layer 3. The FEE in this example is *epäaitoja* 'fake'. The whole noun phrase *epäaitoja koruja* 'fake jewellery'

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

is annotated as ENTITY since the adjective *epäaito* 'fake' is a dependent of the head (*koruja* 'jewellery') of the NP. On the third layer, *koruja* 'jewellery' is marked with the tag REFERENCE_CATEGORY as it also indicates the type relative to which the entity is evaluated. Not shown in Table 2 is the fact that the FinnFN corpus also provides morphosyntactic and dependency annotations on additional layers to make the corpus useful for researching the interface between syntax and semantics.

Annotation of so-called null instantiations is a feature of BFN that was not used in the FinnFN project. Null instantiation is a category for annotating absent semantic constituents or frame elements (Fillmore et al. 2003:245). The term null refers to the fact that a core frame element is absent. Conceptually salient FEs are not always realized as lexical or phrasal material in a sentence, but they have nevertheless been annotated in BFN since they provide relevant insight into the omissibility of semantic material (Ruppenhofer et al. 2010:24–26).

3. COMPOUNDS AND MORPHOLOGY IN FinnFN

In addition to examining the universal nature of the frames, one of the central aims of the FinnFN project has been to develop an annotation system that acknowledges the characteristics of the Finnish language, such as compounding and a more complex morphology. Since the parts of a compound may instantiate different frame elements, it was necessary to address the fact that in general all Finnish compounds are written as a single word, without spacing between free morphemes. In addition, in Finnish some frame elements often show up only as bound morphemes, a typical example being AGENT, only appearing as a verbal suffix. These prominent characteristics of compounding and rich morphology needed to be taken into account when developing the annotation convention for Finnish.

3.1 Compounds in FinnFN

In English, the few compounds that are written as single words are handled in BFN as indivisible units. In example (3), the modifying part *bath* of *bathrobe* is not annotated with the semantic role USE, even though it denotes the use of the garment. We use a typographical convention for annotating an FEE with capital letters, as in BNF, but we also annotate a frame element, e.g. GARMENT, with a subscript in small caps, which is useful when the two overlap as in example (3).

(3) A third ad showed him in a [$_{\text{GARMENT}}$ BATHROBE]. (Clothing)

Compounds consisting of a head noun and a relational modifying part that are not written as a single indivisible word have been annotated separately in BFN, e.g. as in example (4).

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

(4) [MANNER self-]CONFIDENCE (Certainty)

In Finnish, the matter is somewhat more complicated. In Finnish, compounds are practically always written as single indivisible words, and in theory an unlimited number of new compounds can be formed. In FinnFN, we sometimes need to label the parts of a compound with different semantic roles as the part of a compound that modifies the head most often carries a separate semantic meaning. In the FinnFN annotation, we therefore separated the modifying part from the head of the compound and annotated it as illustrated in examples (5) and (6) by the equivalent compounds *riding helmet* and *ratsastuskypärä*. Note the word spacing in English and the lack of spacing in Finnish, which is also reflected in the annotation.

(5) [USE riding] [ACCOUTREMENT HELMET] (Accoutrements)

(6) [USE ratsastus][ACCOUTREMENT KYPÄRÄ] (Accoutrements)
riding.helmet
'riding helmet'

The FinnFN database is not fully uniform in the annotation of compounds and exhibits some variation in what parts are deemed semantically important enough to be labelled separately. Parts are annotated separately in cases where one part of the compound is the FEE and the other is not, similar to the English head noun with a relational modifier as in examples (7) and (8).

(7) [MANNER self-]CONFIDENCE (Certainty)

(8) [MANNER itse]VARMUUS (Certainty)
self.certainty
'self-confidence'

Some lexicalized compounds have been annotated as a whole in Finnish: *aluspaita* lit.: 'undershirt', 'camisole'; or *rintaliivit*, lit.: 'breast vests', 'bra' as in examples (9) and (10).

(9) [MATERIAL puuvillatrikoisen] ALUSPAIDAN (Clothing)
cotton.tricot.ADJ.GEN *camisole.GEN*
'(of) a camisole made of cotton tricot'

(10) Erään suuren liivitehtaan käsityksen mukaan
one.GEN *big.GEN* *bra.factory.GEN* *view.GEN* *according.to*

[DESCRIPTOR pienikupuisten] (A) RINTALIIVIEN kysyntä on
small.cup.ADJ.PL.GEN *A* *bra.PL.GEN* *demand* *has*

taantunut 3,3 prosentilla vuoden 1966 jälkeen. (Clothing)
dwindled 3.3 *percent.ADE* *year.GEN* *1966* *after*
'According to a major bra factory, the demand for bras with small (A-sized) cups has dwindled by 3.3 percent after 1966.'

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

The part of the compound that evokes the frame is annotated as the FEE, and the other relevant parts of the compound are annotated according to their semantic roles. In example (11), only the head is annotated as the FEE.

(11) [MATERIAL nahkaisiin] [USE ratsastus]HOUSUIHIN (Clothing)
leather.ADJ.PL.ILL *riding.pants.ILL*
'into leather riding pants'

The general idea has been to annotate all the relevant and clearly transparent frame elements of the compound. If the compound is opaque, it has been treated as a single unit. However, drawing the line between the two is not always straightforward. Ruppenhofer et al. (2010:41) admit that distinguishing between conventionalized and non-conventionalized compounds is not always clear in English either. Often it was left for the BFN annotator to decide and evaluate.

Finnish compounding resembles German and Swedish compounding more than that of English, so the comparison with German and Swedish FrameNet-related projects regarding how they have dealt with compound annotation is useful. In the German SALSA project, the compounds have been divided into two units and can thus be annotated separately.[11] In the frame Hostile_encounter the German word *Machtkämpfe* (lit.: 'power struggles', 'struggles for power') consists of the FEE *Kämpfe* and the part denoting the TOPIC of the encounter, *Macht*. The compound is structurally very similar in Finnish; the compound *valtataistelut* consists of lexemes with a similar meaning as in German (*valta* 'power', *taistelut* 'struggles, battles').

The Swedish FrameNet project has developed different strategies for annotating compounds.[12] As in Finnish, non-compositional, fully lexicalized compounds are not given any internal analysis in the lexical source database, which treats solid compounds, i.e. single orthographic words, just as it treats single-word items, and does not normally define their formal structure explicitly (Friberg Heppin & Petruck 2014:68). However, in the Swedish FrameNet project compositional compounds are annotated as far as they are compatible with the frame they evoke. The compound parts are annotated separately especially in the cases where one of the parts is the FEE. Usually the whole compound and the compound head are FEEs evoking the same frame. In these cases, the whole Swedish compound is tagged as an FEE, while the modifier is tagged as the appropriate FE, as in example (12).

(12) [[BODY_PART MAG][SJUKDOM]] (Medical_condition) (Swedish)
stomach.illness
'stomach illness'

Like in the Swedish project, FinnFN treats the whole compound as an FEE when it has transparent compositional parts evoking the same frame.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

A common compound type expresses an item and the material it is made of. In example (13) in Swedish illustrating the `Accoutrements` frame, *guldring* 'gold ring' has been separated into FEE (*ring*) and FE denoting MATERIAL (*guld*).

(13) Hon har många [MATERIAL guld]RINGAR
she *has* *many* *gold.ring.PL*

[BODY_LOCATION i båda öronen] (Accoutrements) (Swedish)
in *both* *ear.PL.DEF*
'She has many gold rings in both ears.'

In Finnish, the compound *villasukat* 'wool socks' is also split into head and modifier, each part being annotated separately as in example (14) and only the head acts as the FEE.

(14) [MATERIAL villa]SUKAT (Clothing)
wool.sock.PL
'wool socks'

We note that the Finnish compound annotation is crafted to be similar to that of German and Swedish extending some of the annotation conventions that were available in the BFN.

3.2 Finnish morphology

Agglutination, a central feature in Finnish, presented a challenge to be solved in FinnFN. To deal with morphemes attached to Finnish word stems, most typically possessive and verbal suffixes and comparatives, an annotation convention resembling that of compounding was further developed in which words were split into morphemes if needed. Morphemes have been annotated separately to apply a BFN frame when a core element of the frame would otherwise have been left out and the meaning of an FE was carried by a morpheme, e.g. in the case of the POSSESSOR being marked only by a possessive suffix, see example (15). This convention allowed all the core elements of the frames to be accounted for when annotating Finnish.

Hebrew is another FrameNet language with rich morphology, but the Hebrew FrameNet has not annotated possessives and other bound morphemes as FEs. Instead, the elements have been marked as externally instantiated, i.e. no single word or phrase is marked as representing the frame (Hayoun & Elhadad 2015) following the annotation conventions of the Spanish project, which did not mark semantic roles such as subject markers expressed by bound morphemes. The Finnish project, by contrast, decided to annotate some of the key morphological units because of the agglutinating nature of the language, i.e. the morphology was seen as too important

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

to be omitted: sometimes the FEE, the most essential element evoking a frame, is expressed only by a suffix, which will be discussed later in this section. In Finnish, a suffix may, for example, correspond to an English preposition as the FEE. We also claim that in order to develop FrameNet into a truly universal semantic annotation system, semantic role labeling for agglutinative and polysynthetic languages should be further developed. The assumption that semantic content can only be carried by independent FEEs is rather limiting from a universal point of view, since semantic content can also be carried by bound morphemes as demonstrated by Finnish. If the FrameNet annotations had originally been developed for West Greenlandic, for instance, it would have been essential to account for bound morphemes typical for polysynthetic languages. This is a central point that we wish to emphasize and develop further in the future.

However, in our current project our annotation deals primarily with independent words and phrases: the annotation has been marked on an independent word if one is present, but on a bound morpheme if not. This is illustrated in examples (15) and (16) with regard to both a possessive suffix and a subject marker.

(15) alushousu-[WEARER ni] (Clothing)
underpants-1SG.POSS
vs.
[WEARER minun] alushousu-ni
my *underpants-1SG.POSS*
'my underpants'

(16) Ateriat söi-[INGESTOR mme] keittiön pitkän pöydän
meal.PL *eat.PST-1PL* *kitchen.GEN* *long.GEN* *table-GEN*
ääressä. (Ingestion)
at
vs.
Ateriat [INGESTOR me] söi-mme keittiön pitkän pöydän ääressä.
meal.PL *we* *eat.PST-1PL* *kitchen.GEN* *long.GEN* *table.GEN* *at*
'We ate the meals at the long kitchen table.'

In at least some of the other languages for which a FrameNet currently exists, the agent is visible in the verb form. It seems, however, that in most of these languages it is not possible to leave the overt subject out, and a situation similar to example (16) above will never arise. One of the exceptions is Japanese, where the person is generally unmarked altogether, whereas in Spanish, Italian and Brazilian Portuguese there is often a person marker on the verb. The Japanese FrameNet has dealt with this by marking the agent as a null instantiation similar to the marking of an absent agent in passive sentences in English. In Finnish, where the subject is always visible (often only) on the verb form, it was deemed relevant to mark the FE on the particular morpheme.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

Possessives and comparatives are among the relevant phenomena for some of the frames although they are not core elements. Of these, possessives as affixes can be found only in Finnish and Hebrew of the current FrameNet languages. Comparative and superlative endings, however, are used for indicating degree in other languages as well. In BFN, one can of course claim that the element DEGREE marks only the intensity of the comparison and not the comparison as such, so there is no need to mark the comparative ending *-er* alone. In example (17), the expression of DEGREE of comparison is only visible in the ending and has been left unmarked in BFN.

(17) It was cut in a TRENDIER [BEHAVIOUR style] (`Trendiness`)

The element DEGREE of comparison is a peripheral element in almost every frame of BFN, so omitting its annotation does not leave out core information of the frame. In some individual cases, BFN has marked DEGREE on an entire word denoting comparison, as in example (18), in which the word *older* (earlier in the same sentence) contrasts with the word *newer*.

(18) It is generally desirable to use older textbooks – many of them, alas, out of print – rather than [DEGREE NEWER] [ENTITY ones]. (`Age`)

Note that the word 'more' is often annotated in BFN with DEGREE as in example (19) annotated with the `Obscurity` frame. In Finnish, it is mandatory to use comparative morphemes for expressing the phenomenon represented by the English word 'more', so if the Finnish comparative morphemes were left unannotated, this information would have to be omitted in the Finnish annotation. Comparative morphemes have therefore often been annotated in FinnFN.

(19) [TIME By the 1740s], [ENTITY Stukeley's beliefs] were becoming [DEGREE more] OBSCURE. (`Obscurity`)

Having introduced the annotation of a bound comparative morpheme for DEGREE, we can use this for annotating the comparative ending (*-mpi*) in *ärsyttävämpi* 'more annoying' along with *tavallista* 'than usual' as DEGREE in example (20), while the rest of the word (*ärsyttävä-*) is the FEE.

(20) Mutta sitten eräänä iltana oli Erkki ollut [DEGREE tavallista]
but *then* *one.ESS* *evening.ESS* *had* *Erkki* *been* *usual.PTV*

ÄRSYTTÄVÄ[DEGREE mpi]. (`Stimulus_focus`)
annoying.COMP
'But then one night Erkki had been more annoying than usual.'

Not all Finnish endings, such as enclitic particles like the interrogative particle *-kO*, carry meanings that are relevant for a frame. Polarity or tense do not affect the

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

frames, and for instance the negation verb did not have any particular role or meaning evoking a frame in FrameNet v1.5. Consequently, particles like *-pAs/-hAn* (various meanings; contradiction or mirative, among others) and *-kin* 'too, also' have not been annotated in any frame in FinnFN.

The original BFN database dealt only with lexical units (Fillmore & Baker 2010:338). However, there are also grammatical structures which evoke semantic, pragmatic or interactional frames on their own (Fillmore & Baker 2010:338), so the current BFN deals with much more than FEEs. The present study shows that in Finnish even the FEE itself can be a bound morpheme or a morphosyntactic structure, even though there are only a few examples of this kind in the current corpus. In the initial translations of annotated English examples, each English FE was matched by a Finnish one, so the meanings conveyed solely through a specific grammatical structure or bound morpheme are underrepresented. The following examples (21)–(23) demonstrate three instances in which FEEs might be annotated in grammatical structures or bound morphemes.

(21) a. [Required_situation He] [Degree absolutely] MUST [Required_situation be replaced]. (Required_event)

b. [Required_situation Hänet] ON [Degree ehdottomasti]
him *is* *absolutely*

[Required_situation korva]-TTAVA. (Required_event)
replace-PTCP.PRS.PASS
'He absolutely must be replaced.'

(22) a. [Figure The cat] is ON [Ground the mat]. (Locative_relation)

b. [Figure Kissa] on [Ground mato]-LLA. (Locative_relation)
cat *is* *mat-ADE*
'The cat is on the mat.'

(23) a. [Subset Only one] OUT OF [Group 535] (Partitive)

b. [Subset Vain yksi] 535:STÄ (Partitive)
only *one* *535:ELA*
'only one out of 535'

The annotation system created for annotating parts of compound words makes it possible to mark bound morphemes as FEEs in a similar way. We annotated text examples with the frame `Partitive`, as in example (23b) above. In this frame, the sought meaning in Finnish can only be conveyed through the elative suffix (*-stA*).

Altogether, the instances where the FinnFN corpus currently deviates from the BFN annotation are in the marking of FEs on personal, possessive and comparative markers, and the marking of the FEE on the case suffix in the sole case of the `Partitive` frame. In future development, it might be justified to expand the range of data to include more varied annotations for different types of morphologically

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

encoded linguistic phenomena to the extent that they encode language-independent meanings.

4. DISCUSSION AND FUTURE WORK

Fillmore & Baker (2010:337) mention that it is reasonable to assume that most of the frames are directly comparable between languages. Some frames, for instance the frames describing cooking and commercial transactions, are generally very similar events across cultures, even though differences in details do of course exist. Some frames, on the other hand, are more culture-specific. The main assumption behind frame semantics is that in order to understand a frame one must always have some amount of background information that helps in interpreting the frame (Fillmore & Baker 2010:318). For instance, to understand the weekdays and descriptions about the flow of time, one must have information about how time is seen and counted in Western cultures. Often this kind of culturally specific information is so thoroughly acquired and subconscious that processing it requires a dedicated and conscious cognitive effort. Many of the original frames are notably tied to Western, especially American, culture and its cultural characteristics.

One of our main goals for this research was to test the universality hypothesis of the frames and see how well the original BFN frames and frame elements apply to the Finnish language. In general, the original frame elements defined by BFN turned out to be suitable for annotating Finnish. However, in a few cases the original frame elements were found to be insufficient for describing certain expressions. For example, a phrase such as *ampua laukaus* (lit.: 'shoot a shot') would be a typical example of the frame `Use_firearm` or `Shoot_projectiles` (depending on the context), yet the word *laukaus* 'a shot' (as an event) could not be annotated with any of the predefined frame elements since the meaning of the word does not fit into the description of the FE PROJECTILE: 'The entity whose rapid, dangerous motion is caused by the Agent and/or Firearm'. However, it would be possible to annotate the noun *laukaus* as the FEE. Another more elaborate option, as pointed out by a reviewer, would be to prevent the EVENT element in the abstract frame `Intentionally_affect` from having the status *Core_Unexpressed*. The frame `Use_firearm` inherits from the frame `Intentionally_affect` through the `Using` frame. If the EVENT element was simply *Core* in `Intentionally_affect`, it would be inherited by `Using` and then by `Use_firearm`. In the FinnFN project, elements from other frames were sometimes used if they were deemed suitable and necessary for the annotation of the frame in Finnish. This is natural as the BFN only lists the FEs that have occurred during the annotation of examples with its current frames. In our data, there were only a few words or phrases that could not be annotated as any of the already defined frame elements.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

Other environments, where a suitable way to annotate the data was not immediately obvious, were the various so-called COLORATIVE CONSTRUCTIONS (Jarva & Kytölä 2007, Hamunen 2012), which are rather typical for Finnish. It is a verb structure consisting of a main verb describing the actual act and a companion verb which describes the manner in which the act is performed. The verb describing the manner or 'style' of the act is the one inflected for person and tense. The examples were retrieved from the newspaper corpus using a common verb in infinitive form (*nauraa* 'to laugh', *juosta* 'to run'), and annotated as FEE with MANNER as in (24).

(24) Chubb alkoi äkkiä kiroilla, kun [SELF_MOVER toinen
Chubb start.PST suddenly swear.INF when second

ja kolmas perämies] JUOSTA [MANNER lönkyttivät]
and third mate run.INF lollop.PST.3PL

kajuutoistaan paikalle. (Self_motion)
cabin.PL.ELA.3PL:POSS place.ALL
'Chubb started to swear, as the second and third mate LOLLOPED from their cabins.'

This was done even though BFN normally has verbs in FEs only as part of longer expressions. In Finnish the second verb can be seen as a description of the manner in which the main act is performed. Another possible solution would have been to annotate the description verb as an FEE and leave the main infinitive verb without a tag.

While the frame elements of BFN proved to be rather suitable for annotating Finnish, some of the frames were slightly problematic. We encountered three types of obstacles for finding examples for the English frames in Finnish: (i) lack of suitable search results (caused by the limitations of the automated search method), (ii) linguistic incompatibility, and (iii) cultural specificity of the original frame. The first of the problems was rather easily solved by expanding the automatic search method to allow a round of manual search. There were 14 frames, for which we failed to get useful samples with an automated search method. However, for most of these frames it was easy to find Finnish examples by manually adjusting the searches, and eventually we found examples for all of them. There were two main reasons why we failed to find examples for the 14 frames automatically. First of all, some of the frames describe an event or phenomenon so rare or specific that our corpus simply did not have an example of it. For example the frame `Bond_maturation` uses special vocabulary that could not be found in our primary corpus, but searching the internet eventually yielded an example. Another reason for failing searches is polysemy: for instance, the FEE in the frame `Exchange_currency` is *vaihtaa* 'to exchange' in this case, but also 'replace' or simply 'change (transitive)', as

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

in *vaihtaa vaatteet* 'to change clothes'. The different meanings of *vaihtaa* occur with different frequency, and in our corpus *vaihtaa* in the meaning 'exchange currency' was rather infrequent. It is therefore not surprising that we ended up having automated search results primarily evoking the most frequent frames. Another example of polysemy is the elative suffix in the `Partitive` frame shown in example (23b). The elative suffix is used for many other purposes as well, so manual extraction was needed in order to include Finnish examples for the `Partitive` frame.

The second problem were frames that did not fit into Finnish as well as they do into English due to linguistic incompatibility, i.e. the expressions are so language-specific or idiomatic that translating them into another language is difficult without losing the original frame in the process, as shown in examples (25a) and (26a) with periphrastic translations (25b) and (26b).

(25) a. [DESIRED_GOAL Sleep] EVADED [EXPERIENCER him].
(`Elusive_goal`) (`Self_motion`)

b. Uni KARTTOI häntä (`Unknown frame`)
sleep *avoid.PST* *him*
'Sleep evaded him.'

(26) a. No matter how hard you try, [WINNER they] [FREQUENCY always] BEAT [LOSER you] [COMPETITION to the huts]. (`Beat_opponent`)

b. Vaikka kuinka yrität, he EHTIVÄT aina ensin
even.if *how* *try.2SG* *they* *arrive.in.time.3PL* *always* *first*
mökille. (`Unknown frame`)
cottage.ALL
'Even if you try your best, they are always the first to arrive at the cottage.'

The English examples (25a) and (26a) represent idiomatic expressions that are hard to translate into Finnish with the original frame intact, i.e. the Finnish translation evokes a different frame or, as in example (26b), currently lacks an appropriate frame altogether. Other FrameNet projects, such as the Slovenian one (Lönneker-Rodman 2007a:163), has reported similar cases in which the English frame is not suitable for the structure of the target language.

The third problem was cultural incompatibility of some frames. Going over the data, we only had difficulty finding Finnish correspondences for a few of the frames that had been found useful within the English-speaking cultural sphere. As BFN has developed through locally funded projects, connections between frames and the priorities set by the embedding culture are to be expected. Interestingly, cultural emphasis stands out within certain themes. Examples of this include the various frames describing law and legal proceedings in BFN as in example (27a) with the periphrastic translation (27b).

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

(27) a. [ACCUSED He] was granted [STATUS conditional] BAIL
[JUDGE by Teesside Magistrates]. (Bail_decision)

b. [JUDGE Teessiden tuomarit] myönsivät [ACCUSED hänelle]
Teesside.GEN *magistrate-PL* *grant.PST.3PL* *3SG.ALL*

[STATUS ehdollisen] VAPAUTUKSEN TAKUITA
conditional.ACC *release.ACC* *bail.PL.PART*
VASTAAN. (Bail_decision)
against
'Teesside Magistrates granted him conditional bail.'

The English FEE in the sentence is *bail*, which translates quite awkwardly into Finnish, since there is no one word for this procedure as no corresponding procedure exists within the Finnish judicial system.

In the Japanese FrameNet project, the researchers also concluded that often the problems in annotating new text are not related to culturally specific aspects (Ohara 2012:1562). According to Ohara, usually when a suitable frame did not exist, a new frame was also needed in English. In other words, the problems may actually be related to the approach to semantics in general trying to list all phenomena and their combinations instead of isolating cross-cultural primitives from which the phenomena are generated. The same can be said about the Finnish project, as all the BFN frames had some equivalent in Finnish. Problems related to cultural differences are further described in articles by Bertoldi & Chishman (2011, 2012). A comparison is also found in Friberg Heppin & Friberg (2012) and Friberg Heppin & Toporowska Gronostaj (2014).

FrameNet has recently sparked an interest in the field of natural language processing (Petruck 2013:2), where it has inspired research in semantic parsing and deep semantic analysis. With growing evidence that frame semantics is largely language-independent, FrameNet provides a useful annotation convention for data mining and automatic question answering within and across languages. FrameNet annotations not only provide a framework for information extraction, but also for natural language generation, where data samples offer expressions and utterances that can be reused when generating a message with a specific frame, which has been tested by Dannélls (2010), for example.

5. CONCLUSION

The starting point and the main objective of the FinnFN project was to test the hypothesis of frame universality and to find out if the frames used in BFN would also be evoked in attested Finnish example sentences. As a result, we can conclude that the majority of frames were directly applicable across English and Finnish. Cultural

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

differences, for example, did not pose a major problem. As a by-product of this research project, we have created the first version of FinnFN, documenting attested Finnish example sentences annotated in the same way as English examples in the original BFN project, with some extensions to the annotation convention for bound morphemes.

The main challenges resulted from the original example data and the approach used for finding Finnish examples for the BFN frames. However, the retrieved and annotated sentences served to corroborate the hypothesis that Finnish speakers evoke the same frames when using Finnish.

In the future, it is worthwhile to further develop the annotation conventions to make them even more suitable for central features of the Finnish language, including the agglutinative morphology. Since the main focus of frame semantics is semantic content, the issues that follow from the structural differences of languages can to some extent be considered secondary. The semantic content is clearly very transferable even between non-related languages, as can be seen from other FrameNet projects as well.

It would also be interesting to know whether any Finnish-specific frames were actually left out as a result of the method of data collection applied in the present study. However, some omissions are an inevitable consequence of using a limited set of frames regardless of whether they were created for Finnish or English. As noted before, the frame list of BFN is in no way comprehensive. Correspondingly, any FrameNet project inevitably ends up with similar limitations as the original BFN project for the simple reason that no lexically-oriented language description of finite length can ever be complete. A more systematic annotation of unrestricted Finnish text would shine a light on this aspect of the project. The incompleteness of BFN is not a problem that would pose difficulties for the development of FinnFN, but it does yield information on the types of frames that should be added in future development as has also been noted by, for example, Palmer & Sporleder (2010). This direction is also pursued through full-text annotations at Berkeley and other FrameNet projects.

A possible follow-up project could be the annotation of examples from a large Finnish dictionary. The Swedish FrameNet (SweFN) project aimed to annotate examples from the entire Swedish SALDO dictionary, annotating every SALDO example with a frame they deemed suitable.[13] Therefore the Swedish project also ended up creating many new frames, including scenarios like `Overcoming_misunderstanding` and `Compensating`, along with several others.[14] This should be the next phase in the FinnFN project in order to ensure lexical coverage. Co-operation around the Swedish FrameNet project might prove to be especially fruitful, since the cultural differences should be even smaller than with BFN, and many of the necessary frames would be readily available.

As the semantic information conveyed by a frame can be extracted when an instance of a frame is identified, the FinnFN data is valuable as testing material for

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

data mining and semantic annotation methods when developing automated Finnish information extraction and message generation.

ACKNOWLEDGEMENTS

We are grateful to FIN-CLARIN for supporting the production of this language resource and to the three anonymous *NJL* reviewers and editor for their insightful remarks.

APPENDIX

FrameNets in other languages

German

http://www.coli.uni-saarland.de/projects/salsa/ (accessed 26 October 2015)

Swedish

http://spraakbanken.gu.se/eng/swefn (accessed 26 October 2015)

http://spraakbanken.gu.se/eng/research/swefn/development-version (accessed 26 October 2015)

Japanese

http://jfn.st.hc.keio.ac.jp (accessed 26 October 2015)

Polish

http://www.ramki.uw.edu.pl/en/publikacje.html (accessed 26 October 2015)

French

https://sites.google.com/site/anrasfalda/ (accessed 26 October 2015)

Danish

http://framenet.dk/ (accessed 26 October 2015)

Korean

http://framenet.kaist.ac.kr/framenet/ (accessed 26 October 2015)

Chinese

http://sccfn.sxu.edu.cn/portal-en/home.aspx (accessed 26 October 2015)

Brazilian Portuguese

http://www.ufjf.br/framenetbr/ (accessed 26 October 2015)

Spanish

http://spanishfn.org/ (accessed 26 October 2015)

Slovenian

Birte Lönneker-Rodman, Collin F. Baker & Jisup Hong. 2008. The new FrameNet Desktop: A usage scenario for Slovenian. *Proceedings of ICGL 2008, The First International Conference on Global Interoperability for Language Resources, 9–11 January 2008*, Hong Kong, 147–154.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

http://www.icsi.berkeley.edu/pubs/ai/ICSI_NewFrameNetDesktop08.pdf (accessed 26 October 2015)

Italian

Alessandro Lenci, Martina Johnson & Gabriella Lapesa. 2010. Building an Italian FrameNet through Semi-automatic Corpus Analysis. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, Valetta, Malta. 12–19.

http://lexitron.nectec.or.th/public/LREC-2010_Malta/pdf/313_Paper.pdf (accessed 26 October 2015)

Hebrew

http://www.openu.ac.il/ISCOL2015/downloads/ISCOL2015_submission_30_c_18.pdf (accessed 26 October 2015)

http://www.icsi.berkeley.edu/pubs/ai/HFN.pdf (accessed 26 October 2015)

NOTES

1. http://urn.fi/urn:nbn:fi:lb-2016121201 – Finnish FrameNet frames with annotated example sentences.
2. https://framenet.icsi.berkeley.edu/fndrupal/framenets_in_other_languages
3. http://lrec2016.lrec-conf.org/media/filer_public/2016/08/02/t05_swefn2016lrec_dimitrios.pdf
4. http://www.lrec-conf.org/proceedings/lrec2016/pdf/434_Paper.pdf
5. https://framenet.icsi.berkeley.edu/fndrupal/index.php?q=frameIndex
6. We used the dependency-parsed FrameNet corpus (Baurer, Fürstenau & Rambow 2012) based on FrameNet 1.5.
7. http://urn.fi/urn:nbn:fi:lb-201405276
8. The analyzed corpus can be accessed via https://korp.csc.fi.
9. http://bionlp.utu.fi/
10. http://urn.fi/urn:nbn:fi:lb-2016121301
11. http://www.coli.uni-saarland.de/projects/salsa/gold/page.php?id=s54
12. http://spraakbanken.gu.se/eng/research/swefn/documentation (accessed 10 October 2016).
13. For the SALDO project, see https://spraakbanken.gu.se/eng/home.
14. http://spraakbanken.gu.se/eng/research/swefn/documentation (accessed 10 October 2016).

REFERENCES

Baker, Collin F., Charles J. Fillmore & John B. Lowe. 1998. The Berkeley FrameNet project. In Association for Computational Linguistics (ed.), *Proceedings of the 17th International Conference on Computational Linguistics (Coling 1998)*, 86–90.

Bauer, Daniel, Hagen Fürstenau & Owen Rambow. 2012. The dependency-parsed FrameNet corpus. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey, 3861–3867.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

Bertoldi, Anderson & Rove Luiza de Oliveira Chishman. 2011. The limits of using FrameNet Frames to build a legal ontology. *CEUR Workshop Proceedings 776*, 207–212.

Bertoldi, Anderson & Rove Luiza de Oliveira Chishman. 2012. Developing a frame-based lexicon for the Brazilian legal language: The case of the Criminal_Process Frame. In Monica Palmirani, Ugo Pagallo, Pompeu Casanovas & Giovanni Sartor (eds.), *AI Approaches to the Complexity of Legal Systems: Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents* (Lecture Notes in Computer Science 7639), 256–270.

Bick, Eckhard. 2011. A FrameNet for Danish. *Proceedings of NODALIDA 2011*. Riga, Latvia (NEALT Proceedings Series 11), 34–41. Tartu: Tartu University Library.

Borin, Lars, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj & Dimitrios Kokkinakis. 2010. Swedish FrameNet++. *Swedish Language Technology Conference 2010*. https://svn.spraakdata.gu.se/sb/fnplusplus/pub/sltc2010.pdf (accessed 23 November 2015).

Borin, Lars, Markus Forsberg & Johan Roxendal. 2012. Korp – the corpus infrastructure of Språkbanken. *Proceedings of Eighth International Conference on Language Resources* (LREC 2012), Istanbul, Turkey, 474–478.

Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó & Manfred Pinkal. 2006. The SALSA corpus: A German corpus resource for lexical semantics. *Proceedings of Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 969–974.

Candito, Mandito, Pascal Amsili, Lucie Barque, Farah Benamara, Gaël de Chalendar, Marianne Djemaa, Pauline Haas, Richard Huyghe, Yvette Yannick Mathieu, Philippe Muller, Benoît Sagot & Laure Vieu. 2014. Developing a French FrameNet: Methodology and first results. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland, 1372–1379. http://www.linguist.jussieu.fr/~amsili/papers/asfalda_lrec2014.final.pdf (accessed 23 September 2015).

Dannélls, Dana. 2010. Applying semantic frame theory to automate natural languagetemplate generation from ontology statements. *Proceedings of the 6th International Natural Language Generation Conference 2010*, 179–183.

Duran, Magali Sanches & Sandra Maria Aluísio. 2011. Propbank-Br: A Brazilian Portuguese corpus annotated with semantic role labels. *8th Brazilian Symposium in Information and Human Language Technology*, 164–168.

Fillmore, Charles J. 1976. Frame semantics and the nature of language. *Origins and Evolution of Language and Speech* (Annals of the New York Academy of Sciences 280), 20–32.

Fillmore, Charles J. 1982. Frame semantics. In Linguistics Society of Korea (ed.), *Linguistics in the Morning Calm*, 111–137. Seoul: Hanshin Publishing Company.

Fillmore, Charles J. & Collin F. Baker. 2010. A frames approach to semantic analysis. In Bernd Heine & Heiko Narrog (ed.), *The Oxford Handbook of Linguistic Analysis*, 313–339. New York: Oxford University Press.

Fillmore, Charles J., Christopher R. Johnson & Miriam R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography* 16(3), 235–250.

Fontenelle, Thierry (ed.). 2003. *International Journal of Lexicography* 16(3). [Special issue on FrameNet]

Friberg Heppin, Karin & Håkan Friberg. 2012. Using FrameNet in communicative language teaching. *Proceedings of the XV EURALEX International Congress in Oslo, Norway*, 640–647.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

Friberg Heppin, Karin & Miriam R. L. Petruck. 2014. Encoding of compounds in Swedish FrameNet. *Proceedings of the 10th Workshop on Multiword Expressions (MWE 2014)*, 67–71.

Friberg Heppin, Karin & Maria Toporowska Gronostaj. 2014. Exploiting FrameNet for Swedish mismatch? *Constructions and Frames* 6(1), 52–72.

Hahm, Younggun, Youngsik Kim, Yousung Won, Jongsung Woo, Jiwoo Seo, Jiseong Kim, Seongbae Park, Dosam Hwang & Key-Sun-Choi. 2014. Toward matching the relation instantiation from DBpedia ontology to Wikipedia text: Fusing FrameNet to Korean. *Proceedings of the 10th International Conference on Semantic Systems*, 13–19. Association for Computing Machinery (ACM).

Hamunen, Markus. 2012. Koloratiivirakenne, liike ja tapa [The colorative construction and manner]. In Ilona Herlin & Lari Kotilainen (eds.), *Verbit ja konstruktiot* [Verbs and constructions], 104–140. Helsinki: Suomalaisen Kirjallisuuden Seura.

Hayoun, Avi & Michael Elhadad. 2015. The Hebrew FrameNet project. *Israeli Seminar on Computational Linguistics*. Ra'anana: The Open University of Israel. http://www.openu.ac.il/iscol2015/downloads/ISCOL2015_submission_30_c_18.pdf (accessed 4 June 2017).

Jarva, Vesa & Samu Kytölä. 2007. The Finnish colorative construction and expressivity. *SKY Journal of Linguistics* 20, 235–272.

Lenci, Alessandro, Martina Johnson & Gabriella Lapesa. 2010. Building an Italian FrameNet through Semi-automatic Corpus Analysis. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta. 12–19.

Lindén, Krister, Heidi Haltia, Antti Laine, Juha Luukkonen, Jussi Piitulainen & Niina Väisänen. To appear. TransFrame – Translating frames in the FinnFrameNet Project. *Language Resources and Evaluation*.

Lönneker-Rodman, Birte. 2007a. Beyond syntactic valence: FrameNet markup of example sentences in a Slovenian–German online dictionary. *Proceedings of Fourth International Seminar on Computer Treatment of Slavic and East European Languages (Slovko 2007)*, Bratislava, Slovakia, 152–164.

Lönneker-Rodman, Birte. 2007b. Multilinguality and FrameNet. Technical Report TR-07-001, ICSI, Berkeley, CA.

Meurs, Marie-Jean, Frédéric Duvert, Frédéric Béchet, Fabrice Lefèvre & Renato de Mori. 2008. Semantic frame annotation on the French MEDIA corpus. *Proceedings of the Sixth Language Resources and Evaluation Conference (LREC 2008)*, Marrakech, Marocco, 1014–1018.

Ohara, Kyoko. 2012. Semantic annotations in Japanese FrameNet: Comparing frames in Japanese and English. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey, 1559–1562.

Ohara, Kyoko., Seiko Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito & Shun Ishizaki. 2004. The Japanese FrameNet project: An introduction. *Proceedings of the Workshop on Building Lexical Resources from Semantically Annotated Corpora*, 9–12.

Palmer, Alexis & Caroline Sporleder. 2010. Evaluating FrameNet-style semantic parsing: The role of coverage gaps in FrameNet. *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010): Poster Volume*, 928–936.

Petruck, Miriam R. L. 2013. Introduction. In Mirjam Fried & Kiki Nikiforidou (eds.), *Advances in Frame Semantics* (Benjamins Current Topics 58), 1–12. Amsterdam: John Benjamins.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press

Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson & Jan Scheffczyk. 2010. FrameNet II: Extended theory and practice. https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf (accessed 19 October 2015).

Subirats, Carlos & Miriam R. L. Petruck. 2003. Surprise: Spanish FrameNet. *Proceedings of CIL* 17.

Tonelli, Sara & Emanuele Pianta. 2008. Frame information transfer from English to Italian. *Proceedings of the Sixth Language Resources and Evaluation Conference (LREC 2008)*, Marrakech, Marocco, 2252–2256.

You, Liping & Kaiying Liu. 2005. Building Chinese Framenet database. *Proceedings of the Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2005)*, 301–306.

Zawisławska, Magdalena, Magdalena Derwojedowa & Jadwiga Linde-Usiekniewicz. 2008. A FrameNet for Polish. *Converging Evidence: Proceedings to the Third International Conference of the German Cognitive Linguistics Association (GCLA'08)*, 116–117.

https://doi.org/10.1017/S0332586517000075 Published online by Cambridge University Press